

# Figure 1

1 GCCTCTCTCCATCAGACACCCCAAGGTTCC 30  
-----  
CGGAGAGAGGTTAGTCTGTGGGGTTCCAAGG

31 ATCCGAAGCAGGCGGAGCACCGAACGCACC 60  
-----  
TAGGCTTCGTCCGCCTCGTGGCTTGCGTGG

61 CCGGGGTGGTCAGGGACCCCATCCGTGCT 90  
-----  
GGCCCCACCAGTCCCTGGGGGTAGGCACGA

91 GCCCCCTAGGAGCCCGCGCCTCTCCTCTGC 120  
-----  
CGGGGGATCCTCGGGCGCGGAGAGGAGACG

121 GCCCCGCCTCTCGGGCCGCAACATCGCGCG 150  
-----  
CGGGGCGGAGAGCCCGGCGTTGTAGCGCGC

151 GTTCCTTTAACAGCGCGCTGGCAGGGTGTG 180  
-----  
CAAGGAAATTGTCGCGCGACCGTCCCACAC

181 GGAAGCAGGACCGCGTCCTCCCGCCCCCTC 210  
-----  
CCTTCGTCCTGGCGCAGGAGGGCGGGGGAG

211 CCATCCGAGTTTCAGGTGAATTGGTCACCG 240  
-----  
GGTAGGCTGAAAGTCCACTTAACCAGTGGC

AGGGAGGAGGCCGACACACCACACCTACAC  
241 ----- 271  
TCCCTCCTCCGGCTGTGTGGTGTGGATGTG  
  
TCCCGCGTCCACCTCTCCCTCCCTGCTTCC  
271 ----- 300  
AGGGCGCAGGTGGAGAGGGAGGGACGAAGG  
  
TCTTGGCGGAGGCGGCAGGAACCGAGAGCC  
301 ----- 330  
AGAACCGCCTCCGCCGTCCTTGGCTCTCGG  
  
AGGTCCAGAGCGCCGAGGAGCCGGTCTAGG  
331 ----- 360  
TCCAGGTCTCGCGGCTCCTCGGCCAGATCC  
  
ACGCAGCAGATTGGTTTATCTTGGAAGCTA  
361 ----- 390  
TGCGTCGTCTAACCAAATAGAACCTTCGAT  
  
AAGGGCATTGCTCATCCTGAAGATCAGCTG  
391 ----- 420  
TTCCCGTAACGAGTAGGACTTCTAGTCGAC  
  
                                  START  
ACCATTGACAATCAGCCATGTCATCCAGGC  
421 ----- 450  
TGGTAACTGTTAGTCGGTACAGTAGGTCCG  
  
                                  M  S  S  R  P  
  
CTCTTGAAAGTCCACCTCCTTACAGGCCTG  
451 ----- 480  
GAGAACTTTCAGGTGGAGGAATGTCCGGAC  
  
          L  E  S  P  P  P  Y  R  P  D

ATGAATTCAAACCGAATCATTATGCACCAA  
481 ----- 510  
TACTTAAGTTTGGCTTAGTAATACGTGGTT  
E F K P N H Y A P S  
  
GCAATGACATATATGGTGGAGAGATGCATG  
511 ----- 540  
CGTTACTGTATATACCACCTCTCTACGTAC  
N D I Y G G E M H V  
  
TTCGACCAATGCTCTCTCAGCCAGCCTACT  
541 ----- 570  
AAGCTGGTTACGAGAGAGTCGGTCGGATGA  
R P M L S Q P A Y S  
  
CTTTTTACCCAGAAGATGAAATTCTTCACT  
571 ----- 600  
GAAAAATGGGTGTTCTACTTTAAGAAGTGA  
F Y P E D E I L H F  
  
TCTACAAATGGACCTCTCCTCCAGGAGTGA  
601 ----- 630  
AGATGTTTACCTGGAGAGGAGGTCCTCACT  
Y K W T S P P G V I  
  
TTCGGATCCTGTCTATGCTCATTATTGTGA  
631 ----- 660  
AAGCCTAGGACAGATACGAGTAATAACACT  
R I L S M L I I V M  
  
TGTGCATTGCCATCTTTGCCTGTGTGGCCT  
661 ----- 690  
ACACGTAACGGTAGAAACGGACACACCGGA  
C I A I F A C V A S

CCACGCTTGCCTGGGACAGAGGCTATGGAA  
691 ----- 720  
GGTGCGAACGGACCCTGTCTCCGATACCTT  
T L A W D R G Y G T  
CTTCCCTTTT TAGGAGGTAGTGTAGGCTACC  
721 ----- 750  
GAAGGGAAAATCCTCCATCACATCCGATGG  
S L L G G S V G Y P  
CTTATGGAGGAAGTGGCTTTGGTAGCTACG  
751 ----- 780  
GAATACCTCCTTCACCGAAACCATCGATGC  
Y G G S G F G S Y G  
GAAGTGGCTATGGCTATGGCTATGGTTATG  
781 ----- 810  
CTTCACCGATAACCGATAACCAATAC  
S G Y G Y G Y G Y G  
GCTATGGCTACGGAGGCTATACAGACCCAA  
811 ----- 840  
CGATAACCGATGCCTCCGATATGTCTGGGTT  
Y G Y G G Y T D P R  
GAGCAGCAAAGGGCTTCATGTTGGCCATGG  
841 ----- 870  
CTCGTCGTTTCCCGAAGTACAACCGGTACC  
A A K G F M L A M A  
CTGCCTTTTGTTCATTGCCGCGTTGGTGA  
871 ----- 900  
GACGGAAAACAAAGTAACGGCGCAACCACT  
A F C F I A A L V I

901 TCTTTGTTACCAGTGTTATAAGATCTGAAA 930  
-----  
AGAAACAATGGTCACAATATTCTAGACTTT  
F V T S V I R S E M

931 TGTCCAGAACAAGAAGATACTACTTAAGTG 960  
-----  
ACAGGTCTTGTTCTTCTATGATGAATTCAC  
S R T R R Y Y L S V

961 TGATAATAGTGAGTGCTATCCTGGGCATCA 990  
-----  
ACTATTATCACTCACGATAGGACCCGTAGT  
I I V S A I L G I M

991 TGGTGTTTATTGCCACAATTGTCTATATAA 1020  
-----  
ACCACAAATAACGGTCTTAACAGATATATT  
V F I A T I V Y I M

1021 TGGGAGTGAACCCAACTGCTCAGTCTTCTG 1050  
-----  
ACCCTCACTTGGGTTGACGAGTCAGAAGAC  
G V N P T A Q S S G

1051 GATCTCTATATGGTTCACAAATATATGCCC 1080  
-----  
CTAGAGATATACCAAGTCTTTATATACGGG  
S L Y G S Q I Y A L

1081 TCTGCAACCAATTTTATACACCTGCAGCTA 1110  
-----  
AGACGTTGGTTAAAATATGTGGACGTCGAT  
C N Q F Y T P A A T

CTGGACTCTACGTGGATCAGTATTTGTATC  
1111 ----- 1140  
GACCTGAGATCGACCTAGTCATAAACATAG  
G L Y V D Q Y L Y H  
ACTACTGTGTTGTGGATCCCCAGGAGGCCA  
1141 ----- 1170  
TGATGACACAACACCTAGGGGTCCTCCGGT  
Y C V V D P Q E A I  
TTGCCATTGTACTGGGGTTCATGATTATTG  
1171 ----- 1200  
AACGGTAACATGACCCCAAGTACTAATAAC  
A I V L G F M I I V  
TGGCTTTTGCTTTAATAATTTTCTTTGCTG  
1201 ----- 1230  
ACCGAAAACGAAATTATTAAAAGAAACGAC  
A F A L I I F F A V  
TGAAAACTCGAAGAAAGATGGACAGGTATG  
1231 ----- 1260  
ACTTTTGAGCTTCTTTCTACCTGTCCATAC  
K T R R K M D R Y D  
ACAAGTCCAATATTTTGTGGGACAAGGAAC  
1261 ----- 1290  
TGTTTCAGGTTATAAAACACCCTGTTTCCTTG  
K S N I L W D K E H  
ACATTTATGATGAGCAGCCCCCAATGTCTG  
1291 ----- 1320  
TGTAATACTACTCGTCGGGGGGTTACAGC  
I Y D E Q P P N V E

AGGAGTGGGTAAAAATGTGTCTGCAGGCA  
1321 ----- 1350  
TCCTCACCCAATTTTACACAGACGTCCGT  
E W V K N V S A G T  
  
CACAGGACGTGCCTTCACCCCCATCTGACT  
1351 ----- 1380  
GTGTCCTGCACGGAAGTGGGGGATGACTGA  
Q D V P S P P S D Y  
  
ATGTGGAAAGAGTTGACAGTCCCATGGCAT  
1381 ----- 1410  
ATCACCTTTCTCAACTGTCAGGGTACCGTA  
V E R V D S P M A Y  
  
ACTCTTCCAATGGCAAAGTGAATGACAAGC  
1411 ----- 1440  
TGAGAAGGTTACCGTTTCACTTACTGTTCG  
S S N G K V N D K R  
  
GGTTTTATCCAGAGTCTTCCTATAAATCCA  
1441 ----- 1470  
CCAAAATAGGTCTCAGAAGGATATTTAGGT  
F Y P E S S Y K S T  
  
CGCCGGTTCCTGAAGTGGTTCAGGAGCTTC  
1471 ----- 1500  
GCGGCCAAGGACTTCACCAAGTGGTCAAG  
P V P E V V Q E L P  
  
CATTA ACTTCGCCTGTGGATGACTTCAGGC  
1501 ----- 1530  
GTAATTGAAGCGGACACCTACTCAAGTCCG  
L T S P V D D F R Q

AGCCTCGTTACAGCAGCGGTGGTAACTTTG  
1531 ----- 1560  
TCGGAGCAATGTCGTCGCCACCATTGAAAC  
P R Y S S G G N F E

AGACACCTTCAAAAAGAGCACCTGCAAAGG  
1561 ----- 1590  
TCTGTGGAAGTTTTTCTCGTGGACGTTTCC  
T P S K R A P A K G

GAAGAGCAGGAAGGTCAAAGAGAACAGAGC  
1591 ----- 1620  
CTTCTCGTCCTTCCAGTTTCTCTTGTCTCG  
R A G R S K R T E Q

AAGATCACTATGAGACAGACTACACAACCTG  
1621 ----- 1650  
TTCTAGTGATACTCTGTCTGATGTGTTGAC  
D H Y E T D Y T T G

GCGGCGAGTCCTGTGATGAGCTGGAGGAGG  
1651 ----- 1680  
CGCCGCTCAGGACACTACTCGACCTCCTCC  
G E S C D E L E E D

ACTGGATCAGGGAATATCCACCTATCACTT  
1681 ----- 1710  
TCACCTAGTCCCTTATAGGTGGATAGTGAA  
W I R E Y P P I T S

CAGATCAACAAAGACAACCTGTACAAGAGGA  
1711 ----- 1740  
GTCTAGTTGTTTCTGTGACATGTTCTCCT  
D Q Q R Q L Y K R N

Eukaryote



ATTTTGACACTGGCCTACAGGAATACAAGA  
1741 ----- 1770  
TAAAACTGTGACCGGATGTCCTTATGTTCT  
F D T G L Q E Y K S  
  
GCTTACAATCAGAACTTGATGAGATCAATA  
1771 ----- 1800  
CGAATGTTAGTCTTGAAC TACTCTAGTTAT  
L Q S E L D E I N K  
  
AAGAACTCTCCCGTTTGGATAAAGAATTGG  
1801 ----- 1830  
TTCTTGAGAGGGCAAACCTATTTCTTAACC  
E L S R L D K E L D  
  
ATGACTATAGAGAAGAAAGTGAAGAGTACA  
1831 ----- 1860  
TACTGATATCTCTTCTTTCACTTCTCATGT  
D Y R E E S E E Y M  
  
TGGCTGCTGCTGATGAATACAATAGACTGA  
1861 ----- 1890  
ACCGACGACGACTACTTATGTTATCTGACT  
A A A D E Y N R L K  
  
AGCAAGTGAAGGGATCTGCAGATTACAAAA  
1891 ----- 1920  
TCGTTCACTTCCCTAGACGTCTAATGTTTT  
Q V K G S A D Y K S  
  
GTAAGAAGAATCATTGCAAGCAGTTAAAGA  
1921 ----- 1950  
CATTCTTCTTAGTAACGTTTCGTCAATTTCT  
K K N H C K Q L K S

GCAAATTGTCACACATCAAGAAGATGGTTG  
1951 ----- 1980  
CGTTTAACAGTGTGTAGTTCTTCTACCAAC  
K L S H I K K M V G  
STOP  
GAGACTATGATAGACAGAAAACATAGGAAGG  
1981 ----- 2010  
CTCTGATACTATCTGTCTTTTGTATCTTCC  
D Y D R Q K T  
CTGATGCCAAGTTGTTTGAGAAATTAAGTA  
2011 ----- 2040  
GACTACGGTTCAACAAACTCTTTAATTCAT  
  
TCTGACATCTCTGCAATCTTCTCAGAAGGC  
2041 ----- 2070  
AGACTGTAGAGACGTTAGAAGAGTCTTCCG  
  
AAATGACTTTGGACCATAACCCCGGAAGCC  
2071 ----- 2100  
TTTACTGAAACCTGGTATTGGGGCCTTCGG  
  
AAACCTCTGTGAGCATCACAAAGTTTTGGG  
2101 ----- 2130  
TTTGGAGACACTCGTAGTGTTTCAAACCC  
  
TTGCTTTAACATCATCAGTATTGAAGCATT  
2131 ----- 2160  
AACGAAATTGTAGTAGTCATAACTTCGTAA  
  
TTATAAATCGCTTTTGATAATCAACTGGGC  
2161 ----- 2190  
AATATTTAGCGAAAACCTATTAGTTGACCCG

2191 TGAACAACCTCCAATTAAGGATTTTATGCTT 2220  
-----  
ACTTGTTGAGGTTAATTCCTAAAATACGAA

2221 TAAACATTGGTTCTTGTATTAAGAATGAAA 2250  
-----  
ATTTGTAACCAAGAACATAATTCTTACTTT

2251 TACTGTTTGAGGTTTTTAAGCCTTAAAGGA 2280  
-----  
ATGACAAACTCCAAAAATTCGGAATTCCT

2281 AGGTTCTGGTGTGAACTAACTTTCACACC 2310  
-----  
TCCAAGACCACACTTGATTGAAAGTGTGG

2311 CC 2312  
--  
GG

Figure 2

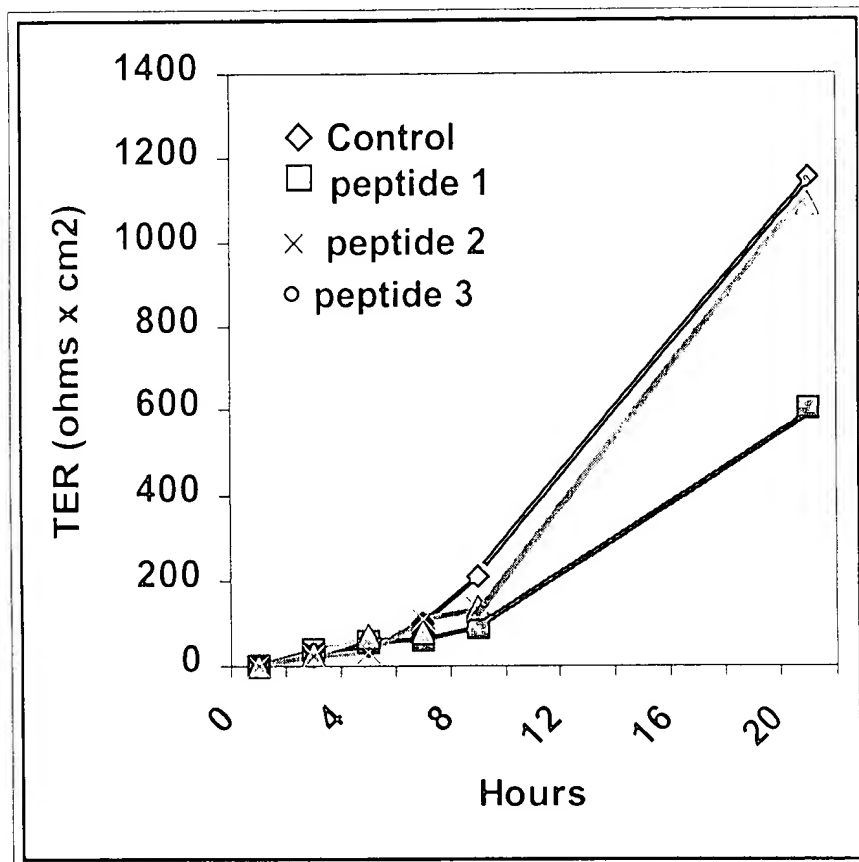


Figure 3A

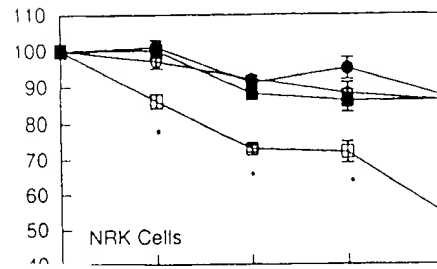


Figure 3B

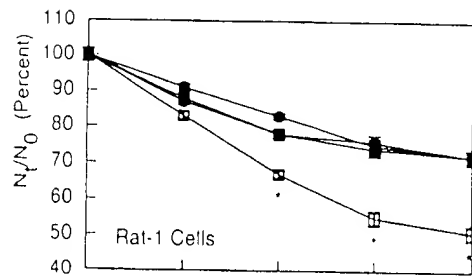


Figure 3C

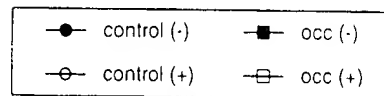
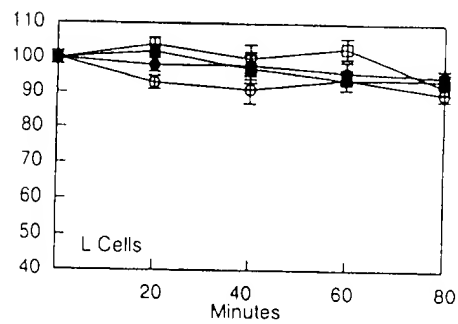


Figure 4

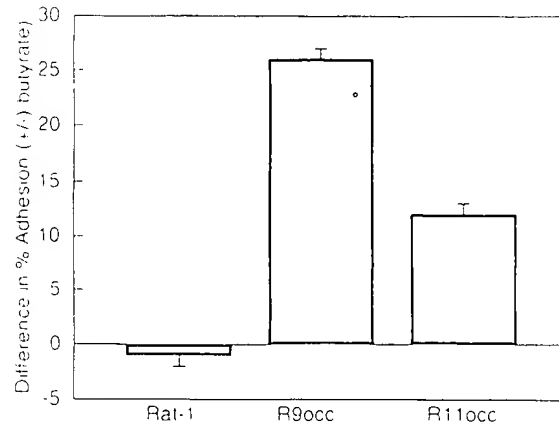


Figure 5A

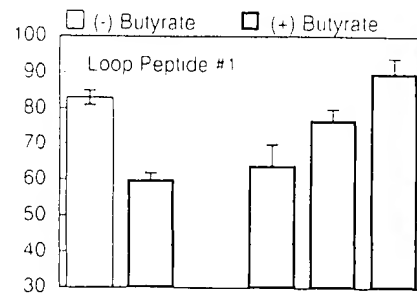


Figure 5B

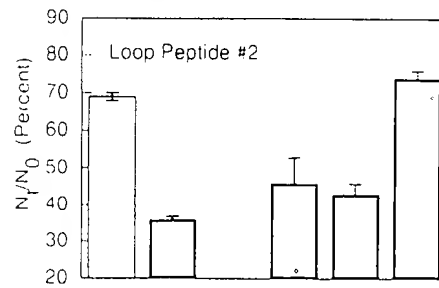


Figure 5C

